Abstract – A survey study about the various methods of feature recognition with machine learning for affective computing is examined. In order to explore the methods of feature recognition with machine learning methods, Sequential Floating Forward Selection (SFFS), Minimum Redundancy – Maximum Relevance (mRMR), Information Gain (IG), and Fisher projection (FP) are discussed. As the machine learning methods, k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) are described. Then, the various feature recognition methods with machine learning methods are compared by the statistical analyses using the classification accuracy performance with applying the discrete emotion data.

Keywords: affective computing, cognitive computing, feature recognition, feature reduction, machine learning

1 Introduction

The research about the automated emotion recognition methods is a very important issue for affective computing and other recognition studies. Picard [1] introduced the study of the affective computing toward machine emotional intelligence with the analysis of affective physiological state. Affective computing is an important study for the development of systems and devices that can recognize, interpret, process, and simulate human affects. In general, the emotion is always interacting with various thinking and affected by the intelligent functioning. LeDoux [2] showed an experimental study using a rat to examine the emotional-oriented processing. Moreover, the brain before stimulated by the incoming signals can be functioning based upon processing the emotion. Hence, the recognition of the emotional senses can be detected by performing pattern recognition of the data based upon the act or emotional senses. In order to recognize the pattern recognition more efficiently, various techniques can be used for recognizing features with intelligent methodologies. Therefore, feature reduction methods and machine learning methods are very useful to develop the study of affective computing. In fact, affective computing fields can be spanning computer science, psychology, and cognitive science that relates to, arises from, or deliberately influences emotion or other affective phenomena.

2 Purpose of Research

The emotion can be formed by the fundamentals of human experiences such as influencing cognition, perception, learning, communication, rational decision-making and etc. Since the range of means and modalities of emotion expression is very wide and various because of the inaccessible factors such as blood chemistry, brain activity, neuro-chemicals, and etc. In order to differentiate the unclassified or unknown factors to evaluate the emotional senses more efficiently, it is very helpful if an automated intelligent method can determine or classify the fuzzy information of the emotional sense data into the crispy and discrete groups or clusters. In addition, among various application fields of studying the classification of the emotional senses, Human-Computer Interaction (HCI) is a very important field to apply and recognize a personalized adaptive emotion to evaluate the emotional response characteristics of individual person and enhance the accuracy of its estimation using appropriate methods.

3 Methods of Research

Through this research, the paper shows a survey of preferable methods with feature reduction and machine learning methods for estimating emotion based on physiological signals. The methods for the feature reduction can be divided into feature selection and feature transformation. For the feature selection, Sequential Floating Forward Selection (SFFS), Minimum Redundancy – Maximum Relevance (mRMR), and Information Gain (IG) are discussed. For the feature transformation, Fisher Projection (FP) is described. For the Machine Learning Methods, k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) are explored.
3.1 Sequential Floating Forward Selection (SFFS)

Sequential Floating Forward Selection (SFFS) was introduced by Pudil et al. [3] and is a widely used methodology in the field of the feature selection. The SFFS consists of a forward (insertion) step and a conditional backward (deletion) step that partially avoids the local optima of the correct classification rate (CCR) [4] with applying the basic Sequential Forward Selection (SFS) procedure starting from the current feature set followed by the successive conditional exclusion of the worst feature in the newly updated set. The procedure of SFFS is following [5]:

Suppose k, the set of features, is determined. Then, apply one step of SFS algorithm followed by the increment of the next features to find the least significant feature in the set. If the current feature is the least significant feature in the set, then, repeat the procedure of SFS. But if the current feature is not the least significant feature, exclude the feature between the previous feature and the current feature. Then, in order to continue to find the least significant feature in the updated set, examine the updated subset and repeat the procedure if the subset is not meeting the conditions.

3.2 Minimum Redundancy Maximum Relevance (mRMR)

The minimum Redundancy-Maximum Relevance (mRMR) approach [6][7] is to select a feature subset, each of which has the maximal relevance with target class and the minimal redundancy with other features based on recognizing that the combinations of individually good variables do not necessarily lead to good classification/prediction performance. The feature subset can be obtained by calculating the mutual information between the features themselves and between the features and the class variables. In other words, to maximize the joint dependency of top ranking variables on the target variable, the redundancy of the feature subset must be reduced, which suggests incrementally selecting the maximally relevant variables while avoiding the redundant ones. The following procedure summarizes mRMR steps. First, the mutual information (MI) between two features \( x \) and \( y \) is evaluated. Then, the average mutual information \( MI(x, l) \) between classification variable \( l \) and feature \( x \) is also calculated. To calculate the potential value of mRMR and achieve feature subset, the two conditions should be optimized, subtract the redundancy by calculating the minimum redundancy condition to minimize the total redundancy of all features selected from the relevance by calculating the maximum relevance condition is to maximize the total relevance between all features in \( S \) and classification variable.

3.3 Information Gain (IG)

Information gain is based on information theory [8] and a univariate method that selects features on the basis of the information contribution related to the class variable without considering feature interactions [13] and works well with texts [10]. It measures the expected reduction in entropy of class before and after observing the features. The information gain of term \( t \) is defined by the \( i^{th} \) category, the probability of the \( i^{th} \) category, and the conditional probabilities that the term \( t \) appears or not in the documents. The information gain indicates the amount of additional information about the features which measured as the entropy using the prior probability for the \( i^{th} \) discretized value of the feature.

3.4 Fisher Projection

The Fischer projection was introduced by Hermann Emil Fischer in 1891 [11], and originally proposed for the representation of carbohydrates. Hence, it is used by chemists, particularly in organic chemistry and biochemistry in order to display a 2-D representation of a 3-D organic molecule by the projection. The use of Fischer projections [12] in non-carbohydrates is discouraged; as such drawings are ambiguous when confused with other types of drawing. The Fischer Projection consists of both horizontal and vertical lines, where the horizontal lines represent the atoms that are pointed toward the viewer while the vertical line represents atoms that are pointed away from the viewer. The point of intersection between the horizontal and vertical lines represents the central element. First convert Dashed-Wedged Line Structure to a “flat” Dashed-Wedged Line Structure to apply the Fischer Projection. The horizontal line represents atoms that are coming out to the front side and the vertical line represents atoms that are going into the backside. The cross image to the right of the arrow is a Fischer projection.

3.5 \( k \)-Nearest Neighbor (\( k \)-NN)

The \( k \)-Nearest Neighbors (\( k \)-NN or \( k \)NN) is a simple classifier which makes use of instance based learning to classify samples [14][15]. The training data is stored and when there is a query on a new sample, the unclassified data is compared with the training data to determine the \( k \)-nearest neighbors. The type of the majority of these \( k \)-nearest neighbors is used to predict the class of the new sample. Even though \( k \)-NN is simple to execute, the run-time performance is
heavily dependent on the amount of training data. The value of k is decided based on the performance of the classifier for a range of k values using 10-fold cross validation method.

### 3.6 Support Vector Machine (SVM)

The SVM classifier uses input features from two classes to determine a maximum margin hyperplane that separates the two classes [16][17]. If the features are not linearly separable, special kernel functions can be used to transform the data to a higher dimensional feature space. Some commonly used kernel functions are linear, polynomial, Radial Basis Function (RBF), and sigmoid kernel functions. Typically, one or more of these kernel functions are applied to the data, and function resulting in the highest accuracy for the least amount of features is selected.

### 3.7 Multilayer Perceptron (MLP)

A multilayer perceptron (MLP) [18] is a feedforward artificial neural network model arranged in layers with forward connections in subsequence layers as a supervised network. The feedforward connections have weights and map the input data onto the appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. A MLP utilizes a supervised learning technique called backpropagation for training the network. A MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable. A graphical representation of an MLP is shown below:

![The structure of the neural network layers for multilayer perceptron](image)

The first layer is the input layer, and the inputs can be distributed into the subsequence layer as the first hidden layer followed by the second hidden layer. Each input sums its inputs, adds a bias or threshold term to the sum and non-linearly transforms the sum to produce an output. The output layer collects all outputs from the second hidden layer for the outputs. Mellit et al. [18] shows the steps of the learning process in the multilayer perception with given a finite input pattern, \( \{x(k)\} \in \mathbb{R} \) where \( 1 \leq k \leq K \).

1. Select the total number of layers, the number of the neurons in each hidden layer, and an error tolerance parameter.
2. Select the initial value of the weight vectors.
3. Initialize the weight vectors.
4. Calculate the neural outputs.
5. Calculate the output error.
6. Calculate the output delta values.
7. Calculate the propagation errors of the hidden neurons from the subsequence layers.
8. Calculate the hidden delta values.
9. Update weight vector.
10. Calculate the error function.
11. If \( k = K \), go to Step 12, otherwise increment \( k \) by 1 and go to Step 4.
12. If the error function is less than or equal to the error tolerance parameter, go to Step 13. Otherwise, go to Step 3.
13. Learning is completed and output the weights.

### 4 Analyses and Results

Using the examined methodologies of feature reduction and machine learning methods, Kukolja et al.[9] shows the comparison of the methodologies using the sample data of the discrete emotion such as 89 samples of sadness, 99 samples of disgust, 38 samples of fear, 78 samples of happiness and 226 samples of other categories which are not suitable for the analysis of the classification of discrete emotions. First of all, the comparison of the classification accuracy using different feature reduction and machine learning methods shows the performance [9] as following on TABLE 1.

Based upon TABLE 1, using the combination of SFFS and machine learning methods, the best classification performance was 60.30 % using SFFS and MLP. Using the combination of mRMR and machine learning methods, the best classification performance was 57.61 % using mRMR and MLP. Using the combination of IG and machine learning methods, the best classification performance was 44.47 % using IG and MLP. Hence, MLP shows the best classification performance without using FP in the feature reduction methods.
TABLE 1  The comparison of feature reduction methods and machine learning methods in the classification accuracy performance using the discrete emotion data [9]

<table>
<thead>
<tr>
<th>Feature reduction methods</th>
<th>Machine learning methods</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kNN (%)</td>
<td>SVM(%)</td>
</tr>
<tr>
<td>SFFS</td>
<td>49.24</td>
<td>50.00</td>
</tr>
<tr>
<td>SFFS-FP</td>
<td>56.18</td>
<td>57.61</td>
</tr>
<tr>
<td>mRMR</td>
<td>50.33</td>
<td>49.13</td>
</tr>
<tr>
<td>mRMR-FP</td>
<td>49.02</td>
<td>48.70</td>
</tr>
<tr>
<td>IG</td>
<td>36.01</td>
<td>41.52</td>
</tr>
<tr>
<td>IG-FP</td>
<td>42.52</td>
<td>41.52</td>
</tr>
<tr>
<td>Average (%)</td>
<td>47.22</td>
<td>48.08</td>
</tr>
</tbody>
</table>

However, applying FP with the feature reduction methods, kNN shows a relatively better performance in the classification performance. In overall, the best average performance using feature reduction methods and machine learning methods is MLP as 49.60 %. In the other hand, the best feature reduction method with the various machine learning methods is SFFS-FP with 55.93 % in the classification accuracy performance. In overall, the total average classification accuracy performance is 48.30 %.

![Fig. 2 Comparison of Classification Accuracy Performance between Feature Reduction Methods and Machine Learning Methods](image)

5 Conclusions

As a conclusion, each feature reduction and machine learning methods are presented based upon their classification performance based upon the classification accuracy applying the examined methods. Based on the survey of the explored studies, a preferred method for real- time estimator adaptation under the certain conditions is Multilayer perceptron (MLP) with the various feature reduction methods followed by kNN. However, the data used for the evaluation of the classification performance is only for the discrete emotion data which used by Kukolja et al. [9]. Therefore, for the future study, more various data need to be explored to find out the best feature reduction method and machine learning method with estimating the best performance of the feature reduction methods and machine learning methods.

6 References


